JOURNAL OF MULTIVARIATE ANALYSIS 4, 52–65 (1974)

Jensen's Inequality for a Convex Vector-Valued Function on an Infinite-Dimensional Space*

MICHAEL D. PERLMAN

University of Chicago, Chicago, Illinois, and
University of Minnesota, Minneapolis, Minnesota

Communicated by M. M. Rao

Jensen's inequality $f(EX) \leqslant Ef(X)$ for the expectation of a convex function of a random variable is extended to a generalized class of convex functions f whose domain and range are subsets of (possibly) infinite-dimensional linear topological spaces. Convexity of f is defined with respect to closed cone partial orderings, or more general binary relations, on the range of f. Two different methods of proof are given, one based on geometric properties of convex sets and the other based on the Strong Law of Large Numbers. Various conditions under which Jensen's inequality becomes strict are studied. The relation between Jensen's inequality and Fatou's Lemma is examined.

1. INTRODUCTION

Jensen's inequality for the expectation of a convex real-valued function of several real variables is as follows:

PROPOSITION 1.1. *Let f be a convex function defined on a convex subset C of n-dimensional Euclidean space R^n, and let $X = (X_1, ..., X_n)$ be an integrable random vector such that $P[X \in C] = 1$. Then $EX \in C$, $Ef(X)$ exists, and*

$$f(EX) \leqslant Ef(X). \tag{1.1}$$

Received February 6, 1973; revised August 1, 1973.

AMS 1970 subject classifications: Primary 28A40; Secondary 60B05.

Key words and phrases: Jensen's inequality, strict inequality, convex vector-valued function, partial ordering, closed cone ordering, linear topological space, dual space, convex set, supporting hyperplane, separating hyperplane, Strong Law of Large Numbers, Banach space, Frechet space, W-convexity, W-continuity.

* Research supported at the University of Chicago by National Science Foundation Grant No. GP-25911 and at the University of Minnesota by U.S. Army Research Office Grant No. DA-ARO-D-31-124-70-G102.

Furthermore, if f is strictly convex and the distribution of X is not concentrated at a single point, then strict inequality holds in (1.1).

This well-known result has many applications in areas such as probability and statistics (cf. [6, 11]). The fact that EX lies in C is proved most easily by induction on the dimension n, while the inequality (1.1) follows from the existence of supporting hyperplanes for a convex set in R^{n+1}; see Ferguson [6, p. 76] for details.

Jensen's inequality can be extended to convex vector-valued functions f: $C \rightarrow R^k$ if the convexity of f is defined with respect to a certain type of partial ordering, called a closed cone ordering (Section 2), on its range in R^k. As an example, let C^p be the class of all $p \times p$ symmetric positive definite matrices. For s_1, s_2 in C^p, define the partial ordering $s_1 \preccurlyeq s_2$ to mean that $s_2 - s_1$ is positive semidefinite. Then the vector-valued function f: $C^p \rightarrow C^p$ defined by $f(s) = s^{-1}$ is convex with respect to this ordering. If S is a random positive definite matrix which is component-wise integrable, then Proposition 1.1 can be applied [7, 16]) to show that $E(S^{-1}) - (ES)^{-1}$ is positive semidefinite, i.e.,

$$f(ES) \preccurlyeq Ef(S), \tag{1.2}$$

a generalized version of (1.1). Here, the partial ordering $\preccurlyeq$ is a closed cone ordering since $\{s \mid 0 \preccurlyeq s\}$, the set of all positive semidefinite $p \times p$ matrices, is a closed convex cone. Notice that the ordering $\preccurlyeq$ is determined by a collection of linear functionals on C^p, since $s_1 \preccurlyeq s_2$ iff $t's_1t \leqslant t's_2t$ for all vectors $t \in R^p$.

Consider now a convex real-valued function f defined on a convex subset C of an infinite-dimensional linear topological space $\mathscr{X}$. Let X be an $\mathscr{X}$-valued random variable defined on a probability space $(\Omega, \mathcal{C}\!\ell, P)$ with range in C, and assume the Pettis integral EX exists (see Section 3). In this case, Jensen's inequality (1.1) no longer holds in general. For example, let $\mathscr{X} = R^\infty$ (the space of all infinite sequences of real numbers with the topology of point-wise convergence), and let C be the set of all nonnegative sequences. Let $\{X_n\}$ be a sequence of nonnegative integrable real-valued random variables defined on $(\Omega, \mathcal{C}\!\ell, P)$ and let X: $\Omega \rightarrow C$ be given by

$$X(\omega) = (X_1(\omega), X_2(\omega), \ldots).$$

Define the real-valued convex function f on C by

$$f(x_1, x_2, \ldots) = \begin{cases} \limsup x_n & \text{if } \limsup x_n < \infty, \\ 0 & \text{if } \limsup x_n = \infty. \end{cases} \tag{1.3}$$

Jensen's inequality (1.1), if it were true in this case, would imply a version of Fatou's Lemma, i.e.,

$$\limsup EX_n \leqslant E(\limsup X_n). \tag{1.4}$$

It is well-known, however, that (1.4) is not true without further assumptions. (Let $(\Omega, \mathcal{A}, P)$ be the unit interval (0, 1) with Lebesgue measure and consider $X_n(\omega) \equiv nI_{(0,n^{-1})}(\omega)$.) In this example, Jensen's inequality fails because the function f, although convex, is nonetheless quite irregular on C. This is in marked contrast to the finite-dimensional case, where a real-valued convex function must be continuous on the interior of its domain. Therefore, to extend Jensen's inequality to an infinite-dimensional space, some continuity assumption must be imposed on the convex function f (see Theorems 3.2–3.10 and 4.1). (The example of this paragraph is discussed further after Theorem 3.6).

In this paper we obtain generalizations of Jensen's inequality of the forms (1.1) and (1.2) for a convex function f whose domain and range are subsets of (possibly) infinite-dimensional linear spaces. Convexity of f is taken with respect to certain closed cone orderings (discussed in Section 2) or more general binary relations (Section 4). The generalized Jensen's inequalities in Section 3 are obtained by utilizing geometrical properties of convex sets in a linear topological space, while in Section 4 a short proof is based on the Strong Law of Large Numbers in a Banach space. This latter provides, incidentally, a very short proof of (1.1) for the finite-dimensional case. Conditions for strict inequality are carefully examined in Section 3. This last topic is complicated by the fact that there are several different ways to define strict inequality with respect to a partial ordering.

The following notations and conventions are used throughout. All linear topological spaces (LTS) considered are understood to be real and Hausdorff. If $\mathcal{X}$ is a LTS, $\mathcal{X}^*$ denotes the dual space of all real-valued continuous linear functionals on $\mathcal{X}$, and $\mathcal{B}(\mathcal{X})$ denotes the Borel σ-field generated by all open subsets of $\mathcal{X}$. The zero element of a LTS is denoted by φ, real scalars by α (with or without subscripts), and the real line by R. The interior, closure, and boundary of a set A are denoted by A^0, $\bar{A}$, and ∂A respectively.

2. Partial Orderings in a Linear Topological Space and Convex Vector-Valued Functions

In this section we discuss the partial orderings used to define convexity of a function f whose range is a subset of a LTS. Let $\mathcal{Y}$ be a LTS and $\preccurlyeq$ a partial ordering on $\mathcal{Y}$, i.e., $\preccurlyeq$ is *reflexive* ($y \preccurlyeq y$) and *transitive* ($y_1 \preccurlyeq y_2 \preccurlyeq y_3 \Rightarrow y_1 \preccurlyeq y_3$). We say $\preccurlyeq$ is a *closed cone ordering* on $\mathcal{Y}$ if it satisfies two additional properties:

(i) $y_1 \preccurlyeq y_2,\ y_3 \in \mathcal{Y},\ 0 \leqslant \alpha \Rightarrow \alpha(y_1 + y_3) \preccurlyeq \alpha(y_2 + y_3)$;

(ii) if $\{y_i\}$ and $\{z_i\}$ are convergent nets in $\mathcal{Y}$ such that $y_i \preccurlyeq z_i$ for all i, then $\lim y_i \preccurlyeq \lim z_i$.

The ordering $\leqslant$ is called *antisymmetric* if $y \leqslant z$, $z \leqslant y \Rightarrow y = z$.

A set $K \subset \mathscr{Y}$ is a *closed convex cone* if K is closed, convex, and if $y \in K$, $\alpha \geqslant 0 \Rightarrow \alpha y \in K$ (hence $\varphi \in K$). We say K is *pointed* if $K \cap (-K) = \{\varphi\}$. Each closed convex cone K determines a closed cone ordering $\leqslant_K$ defined as follows: $y \leqslant_K z$ iff $z - y \in K$. Conversely, if $\leqslant$ is a closed cone ordering, the set $K = \{y \mid y \in \mathscr{Y}, \varphi \leqslant y\}$ is a closed convex cone and the induced ordering $\leqslant_K$ coincides with the original $\leqslant$. Thus, there is a 1–1 correspondence between closed cone orderings and closed convex cones. Furthermore, K is pointed iff $\leqslant_K$ is antisymmetric.

If T is an arbitrary subset of $\mathscr{Y}^*$, the set

$$K_T = \bigcap_{t \in T} \{y \mid t(y) \geqslant 0\}$$

is a closed convex cone and therefore determines a closed cone ordering $\leqslant_T \equiv \leqslant_{K_T}$. We call $\leqslant_T$ the *component-wise ordering* determined by T. The mapping $T \rightarrow K_T$ need not be 1–1, and not every closed cone ordering need be a component-wise ordering. It follows from well-known separation theorems (e.g. [5, Proposition 21.17]), however, that if K is a closed convex cone with $K^0 \neq \varnothing$, then there exists $T \subset \mathscr{Y}^*$ such that $K = K_T$. Furthermore, if $\mathscr{Y}$ is locally convex, then the assumption $K^0 \neq \varnothing$ can be dropped ([5, Corollary 21.15]), so in this case the set of closed cone orderings and the set of component-wise orderings coincide. We say T is *total* on $\mathscr{Y}$ if $t(y) = 0\ \forall t \in T \Leftrightarrow y = \varphi$. T is total iff K_T is pointed. If $\mathscr{Y}$ is locally convex, then $\mathscr{Y}^*$ is total on $\mathscr{Y}$.

For any partial ordering $\leqslant$, we write $y < z$ to indicate that $y \leqslant z$ and $y \neq z$. If $\leqslant = \leqslant_T$ is a component-wise ordering, we write $y \lll z$ to mean that $t(y) < t(z)\ \forall t \in T$. Clearly $y \lll z \Rightarrow y < z$.

Now let C be a convex subset of a LTS $\mathscr{X}$, $f: C \rightarrow \mathscr{Y}$ a vector-valued function, and $\leqslant$ a closed cone ordering on $\mathscr{Y}$. We say f is *convex with respect to* $\leqslant$ if

$$f(\alpha u + (1 - \alpha)v) \leqslant \alpha f(u) + (1 - \alpha) f(v)$$

whenever $u, v \in C$ and $0 \leqslant \alpha \leqslant 1$. It easily follows that for each integer $n \geqslant 2$,

$$f\left(\sum_{k=1}^{n} \alpha_k x_k\right) \leqslant \sum_{k=1}^{n} \alpha_k f(x_k)$$

whenever each $x_k \in C$ and $\alpha_k \geqslant 0$ with $\sum \alpha_k = 1$. The function f is *strictly convex with respect to* $\leqslant$ if

$$f(\alpha u + (1 - \alpha)v) < \alpha f(u) + (1 - \alpha) f(v)$$

whenever $u, v \in C$, $u \neq v$, $0 < \alpha < 1$. If $\preccurlyeq$ is in fact a component-wise ordering, f is *strictly component-wise convex with respect to* $\preccurlyeq$ if

$$f(\alpha u + (1-\alpha)v) \lll \alpha f(u) + (1-\alpha) f(v)$$

whenever $u, v \in C$, $u \neq v$, $0 < \alpha < 1$.

3. Jensen's Inequality for Pettis Integrable Functions

The generalizations of Jensen's inequality obtained in this section are valid under the minimal assumption of Pettis integrability. We take $\mathscr{X}$ to be a LTS and $(\Omega, \mathcal{A}, P)$ a probability space. A mapping $X: \Omega \to \mathscr{X}$ is *Pettis integrable* [9, 14, 15] with respect to $(\Omega, \mathcal{A}, P)$ if (a) X is *weakly measurable*, i.e., $x^*(X)$ is $\mathcal{A}$-measurable $\forall x^* \in \mathscr{X}^*$; (b) $\int x^*(X)\, dP$ exists $\forall x^* \in \mathscr{X}^*$; and (c) there exists $\hat{x}_A \in \mathscr{X}$ such that $x^*(\hat{x}_A) = \int_A x^*(X)\, dP$ $\forall x^* \in \mathscr{X}^*$, and $A \in \mathcal{A}$. We denote any such element $\hat{x}_A$ by $(EX)_A$, which may not be uniquely determined. If $\mathscr{X}^*$ is total on $\mathscr{X}$, however, then the Pettis integral $(EX)_A$ is uniquely determined if it exists; in particular, this is the case if $\mathscr{X}$ is locally convex. In this paper, we only use the weaker condition in (c) that $\hat{x}_A \in \mathscr{X}$ exists for one $A = \Omega$ and write $(EX)_\Omega$ simply as EX.

A closed hyperplane $H = \{x \mid x^*(x) = \alpha\}$ is a *supporting hyperplane* of a set $A \subset \mathscr{X}$ if $x^*(A) \leqslant \alpha$ (or $x^*(A) \geqslant \alpha$) and $H \cap A \neq \varnothing$. Theorem 3.1 is an easy extension of a result in Bourbaki ([3, Théorème 1, Chap. 4, Section 6]).

Theorem 3.1. *Suppose X is Pettis integrable. Let C be a convex subset of $\mathscr{X}$ such that the range $X(\Omega) \subseteq C$.*

(i) *If either* (a) $C^0 \neq \varnothing$ *or* (b) *$\mathscr{X}$ is locally convex, then $EX \in \bar{C}$.*

(ii) *If $C^0 \neq \varnothing$ and if $P[x^*(X) = \alpha] < 1$ whenever $\{x \mid x^*(x) = \alpha\}$ is a supporting hyperplane of $\bar{C}$, then $EX \in C^0$.*

Proof. (i) If either (a) or (b) hold, then $\bar{C}$ can be expressed as an intersection of closed halfspaces, i.e., $\bar{C} = \bigcap_i \{x \mid x_i^*(x) \leqslant \alpha_i\}$ ([5, Corollary 21.14 and Proposition 21.17]). For each i, $P[x_i^*(X) \leqslant \alpha_i] = 1$ since $X(\Omega) \subseteq C$, so $\alpha_i \geqslant Ex_i^*(X) = x_i^*(EX)$, hence $EX \in \bar{C}$.

(ii) Suppose $EX \in \partial C$. Since $\bar{C}$ is a convex body, we can assume that $x_i^*(EX) = \alpha_i$ for some i ([5, Proposition 21.17]). However, the hypothesis implies that $x_i^*(EX) < \alpha_i$, a contradiction, so $EX \in C^0$.

Remark 3.1. If $\{\omega \mid X(\omega) \in C^0\}$ contains a set of positive P-measure, then the hypotheses of Theorem 3.1(ii) are satisfied, since $C^0 \cap H = \varnothing$ for every supporting hyperplane H of $\bar{C}$.

Remark 3.2. In Theorem 3.1(i) it can happen that $EX \notin C$. For example, let $\mathscr{X} = R^\infty$, and let C be the subset consisting of all sequences having only finitely many nonzero elements. Then $\mathscr{X}$ is locally convex and C is convex. Let U and N be independent real random variables such that $EU = 1$, $P[N = n] > 0$ for $n = 1, 2, \ldots$, and $\sum_{n=1}^{\infty} P[N = n] = 1$. Let $X(n, u)$ be the sequence $(u, \ldots, u, 0, \ldots)$ consisting of n u's and the remainder 0's, so $X(n, u) \in C$. Then $EX(N, U) = (q_1, q_2, \ldots)$ where $q_k = P[N \geqslant k] > 0$, so $EX \notin C$.

We will make use of the following assumptions in Theorems 3.2–3.9:

A.1. $\mathscr{X}$ and $\mathscr{Y}$ are real, Hausdorff LTS.

A.2. C is a convex subset of $\mathscr{X}$.

A.3. $(\Omega, \mathscr{A}, P)$ is a probability space.

A.4. $X\colon \Omega \to \mathscr{X}$ is a Pettis integrable mapping such that $X(\Omega) \subseteq C$.

A.5. $\leqslant_T$ is a component-wise ordering on $\mathscr{Y}$ determined by $T \subset \mathscr{Y}^*$.

A.6. $f\colon C \to \mathscr{Y}$ is convex with respect to $\leqslant_T$.

A.7. $f(X)\colon \Omega \to \mathscr{Y}$ is Pettis integrable.

Jensen's inequality and conditions for strict inequality are obtained first under the assumption that $C^0 \neq \varnothing$, which guarantees the existence of supporting hyperplanes. Later, this assumption is replaced by the assumption that $\mathscr{X}$ is locally convex.

THEOREM 3.2. *Under assumptions* A.1–A.7, *if* $C^0 \neq \varnothing$, *if* $P[x^*(X) = \alpha] < 1$ *for each supporting hyperplane* $\{x \mid x^*(x) = \alpha\}$ *of* $\bar{C}$, *and if* $t(f(\cdot))$ *is continuous on* C^0 *for each* $t \in T$, *then* $f(EX) \leqslant_T Ef(X)$.

Proof. Choose any $t \in T$, and set $h(x) = t(f(x))$. Since $t(Ef(X)) = Et(f(X))$, it suffices to show that $h(EX) \leqslant Eh(X)$. This will be seen to be a consequence of the following representation for the convex function h continuous on C^0: for each x_0 in C^0,

$$h(x_0) = \max\{m(x_0) \mid m \text{ affine}, m \leqslant h \text{ on } C\}, \tag{3.1}$$

where an *affine* function on $\mathscr{X}$ is of the form $m(x) = x^*(x) + \alpha$ for some $x^* \in \mathscr{X}^*$ and $\alpha \in R$.

To obtain (3.1), consider the subset Q of the product space $R \times \mathscr{X}$ given by

$$Q = \{(\alpha, x) \mid \alpha \geqslant h(x), x \in C\}.$$

Q is a convex set since h is convex, and $Q^0 \neq \varnothing$ since h is continuous on C^0. Fix a point $x_0 \in C^0$. Since $(h(x_0), x_0) \in \partial Q$, there exists a supporting hyperplane for Q through this point, i.e., there is a nonzero $g \in (R \times \mathscr{X})^*$ and $c \in R$ such that

$$g(h(x_0), x_0) = c \tag{3.2}$$

and

$$g(\alpha, x) \geqslant c \qquad \text{if} \quad (\alpha, x) \in Q \tag{3.3}$$

([5, Theorem 21.11 and Corollary]). Write $g(\alpha, x) = \alpha g(1, \varphi) + g(0, x)$; we claim that $g(1, \varphi) > 0$. If $g(1, \varphi) < 0$, then $g(\alpha, x) \to -\infty$ as $\alpha \to +\infty$, contradicting (3.3). If $g(1, \varphi) = 0$, then $g(\alpha, x) = g(0, x)$, hence $g(\alpha, x_0) = c$ for all $\alpha \in R$. Since g is nonzero, there is an $x_1 \neq \varphi$ such that $g(0, x_1) = 1$, so

$$g(\alpha, x_0 - \delta x_1) = c - \delta \tag{3.4}$$

for all α, $\delta \in R$. Since $x_0 \in C^0$, $\delta > 0$ can be chosen small enough so that $x_0 - \delta x_1 \in C$, so (3.4) contradicts (3.3). Therefore, $g(1, \varphi) > 0$ as claimed, so we can define

$$m(x) \equiv g(1, \varphi)^{-1}[c - g(0, x)].$$

The function m is an affine function on $\mathscr{X}$ such that $m \leqslant h$ on C (by (3.3)) and $m(x_0) = h(x_0)$ (by (3.2)), which proves (3.1).

Now, to show that $h(EX) \leqslant Eh(X)$, by (3.1) we can choose an affine function $\bar{m}$ such that $\bar{m} \leqslant h$ on C and $\bar{m}(EX) = h(EX)$ ($EX \in C^0$ by Theorem 3.1(ii)). Since X is Pettis integrable, $\bar{m}(EX) = E\bar{m}(X) \leqslant Eh(X)$ as required. (The idea of this proof is similar to one of Meyer ([12, Chap. 11, p. 223]) whose result is stated for compact C.) ∎

If X is not a degenerate (constant) random vector, and if f is strictly component-wise convex, then strict component-wise inequality holds in Jensen's inequality. The inner measure induced by P is denoted by P_*.

THEOREM 3.3. *In addition to the hypotheses of Theorem* 3.2, *assume that f is strictly component-wise convex on C and that $P_*[A(EX)] < 1$, where $A(x) = \{\omega \mid X(\omega) = x\}$. Then $f(EX) \lll Ef(X)$.*

Proof. Using the notation of the preceding proof, we must show that $h(EX) < Eh(X)$. Since h is now strictly convex and the graph of the affine function $\bar{m}(x)$ is a linear surface (i.e., hyperplane), $\bar{m}(x)$ and $h(x)$ can agree at only one point of C. Since $\bar{m}(EX) = h(EX)$, it follows that $\bar{m}(x) < h(x)$ if $x \in C$ and $x \neq EX$, so

$$\Omega \backslash A(EX) = \{\omega \mid \bar{m}(X(\omega)) < h(X(\omega))\}.$$

Therefore, $A(EX) \in \mathcal{Ol}$ since $\bar{m} \circ X$ and $h \circ X$ are measurable, so $P[A(EX)] < 1$. Hence, $P[\bar{m}(X) < h(X)] > 0$, so $Eh(X) > E\bar{m}(X) = \bar{m}(EX) = h(EX)$. ∎

To prove that strict inequality occurs in Jensen's inequality if f is strictly convex with respect to $\preccurlyeq_T$, it must be assumed that T is total on $\mathcal{Y}$, and additional structure must be imposed on either $\mathcal{X}$ or $\mathcal{Y}$. In Theorem 3.4, it is assumed that $\mathcal{Y}$ is a normed linear space, while in Theorem 3.5 it is assumed that $\mathcal{X}$ is a separable pre-Frechet space (a locally convex, metrizable LTS), and additional regularity for f is required.

THEOREM 3.4. *In addition to the hypotheses of Theorem* 3.2, *assume that f is strictly convex with respect to $\preccurlyeq_T$ on C, that $P_*[A(EX)] < 1$, that $\mathcal{Y}$ is a normed linear space, and that there exists a countable subset $T_0 \subset T$ such that T_0 is total on $\mathcal{Y}$. Then $f(EX) \prec Ef(X)$.*

Proof. Let $T_0 = \{t_n\}$, and let $S_0 = \{s_n\} = \{t_n/\| t_n \|\}$, so S_0 is also total on $\mathcal{Y}$ and each $\| s_n \| = 1$. If we define the functional s on $\mathcal{Y}$ by $s(y) = \sum_{n=1}^{\infty} 2^{-n} s_n(y)$, then $s \in \mathcal{Y}^*$, and $s(y) < s(z)$ if $y \prec z$ since S_0 is total. Thus, $h(x) = s(f(x))$ is a strictly convex real-valued function on C, so the argument used to prove Theorem 3.3 shows that $s(f(EX)) \equiv h(EX) < Eh(X) = s(Ef(X))$. Thus $f(EX) \neq Ef(X)$, so $f(EX) \prec Ef(X)$. ∎

Remark 3.3. If T is total on $\mathcal{Y}$, and if $\mathcal{Y}^*$ is separable in the strong (norm) or weak* topology, then the last assumption in Theorem 3.4 is necessarily satisfied (e.g., $\mathcal{Y} = L_p$, $1 < p < \infty$, or $\mathcal{Y}$ a separable Hilbert space). This remark also applies to Theorem 3.8.

THEOREM 3.5. *In addition to the hypotheses of Theorem* 3.2, *assume that f is strictly convex with respect to $\preccurlyeq_T$ on C, that $P_*[A(EX)] < 1$, that $\mathcal{X}$ is a separable metrizable locally convex* LTS, *that $t(f(\cdot))$ is lower semicontinuous on C for each $t \in T$, and that T is total on $\mathcal{Y}$. Then $f(EX) \prec Ef(X)$.*

Proof. First, because of the additional structure assumed for $\mathcal{X}$, the weakly measurable mapping X is in fact a Borel measurable mapping from $(\Omega, \mathcal{Ol})$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. This is proved, for example, by Ahmad ([1, Corollary to Proposition 1, p. 100]) and is an extension of a well-known result of Pettis [14] for a normed linear space (see also [9, Theorem 3.5.3]). In particular, $A(x) \in \mathcal{Ol}$ for each x.

Let $\mathcal{U}$ be a countable collection of open sets forming a base for the topology of $\mathcal{X}$. There must exist some $x_0 \in (C\backslash\{EX\})$ such that $P[X \in U] > 0$ for every open neighborhood U of x_0. For if not, then for each $x \in (C\backslash\{EX\})$, there exists some $U(x) \in \mathcal{U}$ such that $x \in U(x)$ and $P[X \in U(x)] = 0$, implying that the event $\{X \neq EX\}$ can be expressed as a countable union of null events, so

$P[A(EX)] = 1$, a contradiction. Next, since f is strictly convex and T is total on $\mathscr{Y}$, there exists $t \in T$ such that

$$h((1/2)\,x_0 + (1/2)\,EX) < (1/2)\,h(x_0) + (1/2)\,h(EX),$$

where $h(x) = t(f(x))$. Therefore, $\bar{m}(x_0) < h(x_0)$, where $\bar{m}(x)$ is the affine function obtained in the proof of Theorem 3.2 which supports h from below and for which $\bar{m}(EX) = h(EX)$. Since h is lower semicontinuous and $\bar{m}(x)$ is continuous on C, the set $U_0 = \{x \mid x \in C, \bar{m}(x) < h(x)\}$ is relatively open in C and contains x_0, so $P[X \in U_0] > 0$. Therefore, $Eh(X) > E\bar{m}(X) = h(EX)$, implying $t(Ef(X)) > t(f(EX))$, so $f(EX) \prec Ef(X)$. ∎

Remark 3.4. The assumption that $\mathscr{X}$ is locally convex is imposed in Theorem 3.5 only to guarantee (along with separability and metrizability) that the weakly measurable mapping X is in fact Borel measurable, and thus can be dropped if this latter fact is assumed.

Up to now, the assumption $C^0 \neq \varnothing$ has been used to guarantee the existence of supporting hyperplanes for the set $Q = \{(\alpha, x) \mid \alpha \geqslant h(x), x \in C\}$ defined in the proof of Theorem 3.2. For the rest of this section, we drop this assumption and instead assume that $\mathscr{X}$ is locally convex and C is closed. Jensen's inequality is now obtained by using the fact that in a locally convex space two closed convex disjoint sets, one of which is compact, can be strictly separated by a closed hyperplane ([5, Theorem 21.12]).

THEOREM 3.6. *Under assumptions* A.1–A.7, *if C is closed, if $\mathscr{X}$ is locally convex, and if $t(f(\cdot))$ is lower semicontinuous on C for each $t \in T$, then $f(EX) \preccurlyeq Ef(X)$.*

Proof. Choose $t \in T$, and let $h(x) = t(f(x))$. As a consequence of the strict separation property mentioned above, it can be shown by an argument similar to that leading to (3.1) that for each x in C,

$$h(x) = \sup\{m(x) \mid m \text{ affine}, m < h \text{ on } C\}$$

(cf. [5, Proposition 21.18]). For any such m, $m(EX) = Em(X) < Eh(X)$ so $h(EX) = \sup_{m<h} m(EX) \leqslant Eh(X)$, completing the proof (note that $EX \in C$ by Theorem 3.1(i)). ∎

The necessity of the assumption of lower semicontinuity in Theorem 3.6 is illustrated by the example in the third paragraph of Section 1, involving Fatou's Lemma. In the example, $\mathscr{X} \equiv R^\infty$ and the dual space $\mathscr{X}^*$ consists of all linear functionals x^* of the form

$$x^*(x_1, x_2, \ldots) = \sum_{i=1}^{r} \alpha_i x_i$$

for some $r \geqslant 1$ and $\alpha_i \in R$. Thus, $X = (X_1, X_2, ...)$ is Pettis integrable if and only if each X_n is integrable and

$$EX = (EX_1, EX_2, ...).$$

If we take $\mathscr{Y} = R$, let $\preccurlyeq$ be the natural ordering on R, and assume $E(\limsup X_n) < \infty$ (otherwise (1.4) is trivially satisfied); then assumptions A.1–A.7 are satisfied, where f is given by (1.3). Furthermore, C is closed and $\mathscr{X} = R^{\alpha}$ is locally convex, so all assumptions of Theorem 3.6 are satisfied except that f is *not* lower semicontinuous on C (this amounts to the fact that the value of a double limit may change if the order of the two limit operations is reversed). However, as pointed out in Section 1, the conclusion of Theorem 3.6, i.e., inequality (1.4), fails.

Results concerning strict inequality, analogous to Theorems 3.3–3.5, are now presented. The inner measure induced by P is denoted by P_*.

THEOREM 3.7. *In addition to the hypotheses of Theorem* 3.6, *assume that f is strictly component-wise convex with respect to $\preccurlyeq_T$ on C and that $P_*[A(EX)] < 1$. Then $f(EX) \preccurlyeq Ef(X)$.*

Proof. Using the notation of the preceding proof, there exists a sequence $\{m_n\}$ of affine functions such that $m_n < h$ on C and $m_n(EX) \to h(EX)$. If $h(EX) = Eh(X)$, then $E \mid h(X) - m_n(X)| \to 0$, so there exists a subsequence $\{m_k\} \subseteq \{m_n\}$ and a subset $\Omega_0 \in \mathscr{O}$ such that $m_k(X(\omega)) \to h(X(\omega))$ for each $\omega \in \Omega_0$ and such that $P(\Omega_0) = 1$. Now, $X(\Omega_0)$ must contain at least two distinct points, say, x_1 and x_2 (otherwise, if $X(\Omega_0) = \{x\}$ alone, then $x = EX$ and $P_*[A(EX)] = 1$), so $m_k(x_i) \to h(x_i)$, $i = 1, 2$. Therefore,

$$m_k((1/2)\, x_1 + (1/2)\, x_2) \to (1/2)\, h(x_1) + (1/2)\, h(x_2) > h((1/2)\, x_1 + (1/2)\, x_2)$$

(h is strictly convex). Since $m_k < h$, this is a contradiction, so it must be that $h(EX) < Eh(X)$. ∎

To obtain strict inequality when f is strictly convex, we again have to impose further regularity and countability restrictions.

THEOREM 3.8. *In addition to the hypotheses of Theorem* 3.6, *assume that f is strictly convex with respect to $\preccurlyeq_T$ on C, that $P_*[A(EX)] < 1$, that $\mathscr{Y}$ is a normed linear space, and that there exists a countable subset $T_0 \subset T$ such that T_0 is total on $\mathscr{Y}$. Then $f(EX) \prec Ef(X)$.*

Proof. Use the proof of Theorem 3.4 with "Theorem 3.7" substituted for "Theorem 3.3" ∎

The next theorem is quite similar to Theorem 3.5, differing mainly in that C is assumed closed rather then $C^0 \neq \varnothing$.

THEOREM 3.9. *In addition to the hypotheses of Theorem* 3.6, *assume that f is strictly convex with respect to $\leqslant_T$ on C, that $P_*[A(EX)] < 1$, that $\mathscr{X}$ is a separable metrizable locally convex* LTS, *that $t(f(\cdot))$ is continuous on C for each $t \in T$, and that T is total on $\mathscr{Y}$. Then $f(EX) \prec Ef(X)$.*

Proof. As pointed out in the proof of Theorem 3.5, X is Borel measurable. Let x_0 and h be as defined in the proof of Theorem 3.5, and suppose $h(EX) = Eh(X)$. As in the proof of Theorem 3.7, there exists a sequence $\{m_k\}$ of affine functions and a subset $\Omega_0 \in \mathscr{A}$ such that $m_k(EX) \to h(EX)$, $m_k < h$ on C, $m_k(X(\omega)) \to h(X(\omega))$ for each $\omega \in \Omega_0$, and $P(\Omega_0) = 1$. Let $\{U_n\}$ be a decreasing sequence of open neighborhoods of x_0 such that $\bigcap U_n = \{x_0\}$, let ω_n be any element of $X^{-1}(U_n) \cap \Omega_0$ (which is nonempty since $P[X \in U_n] > 0$), and let $x_n = X(\omega_n)$. Then $x_n \to x_0$ and $\lim_{k\to\infty} m_k(x_n) = h(x_n)$ for each n. Since h is continuous on C and

$$h((1/2)\,x_0 + (1/2)\,EX) < (1/2)\,h(x_0) + (1/2)\,h(EX),$$

there exists n' such that

$$h((1/2)\,x_{n'} + (1/2)\,EX) < (1/2)\,h(x_{n'}) + (1/2)\,h(EX).$$

But this is impossible since the affine functions m_k converge to h at each of the three points $x_{n'}$, EX, and $(1/2)\,x_{n'} + (1/2)\,EX$. Thus, $h(EX) < Eh(X)$, so $f(EX) \neq Ef(X)$, hence $f(EX) \prec Ef(X)$. ∎

Conditions under which Jensen's inequality holds for a convex real-valued function f defined on a LTS $\mathscr{X}$ can be obtained simply by taking $\mathscr{Y} = R$ in Theorems 3.2 and 3.6. The following theorem extends these results to the important case of a function which may assume infinite values.

THEOREM 3.10. *Let* A.1–A.4 *be satisfied and let $f = C \to [-\infty, \infty)$ be a convex extended-real-valued function such that $Ef(X)$ exists. If either*

(i) *$C^0 \neq \varnothing$, $P[x^*(X) = \alpha] < 1$ for each supporting hyperplane $\{x \mid x^*(x) = \alpha\}$ of $\bar{C}$, f continuous on C^0, or*

(ii) *C closed, $\mathscr{X}$ locally convex, f lower semicontinuous on C,*

then $f(EX) \leqslant Ef(X)$.

Proof. If $Ef(X) = +\infty$, the desired inequality is trivial, so assume $Ef(X) < \infty$. Under condition (i) (respectively, condition (ii)) for each $\alpha \in R$, the function $\max(f(x), \alpha)$ is also convex and continuous on C^0 (respectively, lower semicontinuous on C), so

$$\max(f(EX), \alpha) \leqslant E \max(f(X), \alpha)$$

by Theorem 3.2 (respectively, Theorem 3.6). The desired inequality follows by letting $\alpha \to -\infty$ and applying the Monotone Convergence Theorem. ∎

The results in this section have been obtained under the assumption that X is Pettis integrable. Some recent results concerning the existence of the Pettis integral ($\equiv$ barycenter) are given by Bourgin [4] and Khurana [10]. In the next section, the stronger assumption that X is Bochner integrable is imposed, necessitating consideration of LTS $\mathscr{X}$ which are metrizable and complete.

4. Jensen's Inequality for Bochner Integrable Functions

In the preceding section the supporting and separating hyperplane theorems were used to obtain Jensen's inequality for Pettis integrable functions convex with respect to a component-wise ordering on $\mathscr{Y}$. When the stronger assumption of Bochner integrability is imposed, however, the Strong Law of Large Numbers (SLLN) easily yields a generalized Jensen's inequality for a wider class of functions, those convex with respect to a more general binary relation on $\mathscr{Y}$, not necessarily a partial ordering. Since a limiting process is used, this method does not apply to questions of strict inequality.

We now impose seven assumptions, corresponding to A.1–A.7 in Section 3:

B.1. $\mathscr{X}$ and $\mathscr{Y}$ are real separable Banach spaces.

B.2. C is a convex subset of $\mathscr{X}$.

B.3. $(\Omega, \mathscr{A}, P)$ is a probability space.

B.4. $X: \Omega \to \mathscr{X}$ is a Bochner integrable mapping such that $X(\Omega) \subseteq C$.

B.5. W is a subset of the product space $\mathscr{Y} \times \mathscr{Y}$.

B.6. $f: C \to \mathscr{Y}$ is W-convex on C (defined below).

B.7. $f(X): \Omega \to \mathscr{Y}$ is Bochner integrable.

A mapping $X: \Omega \to \mathscr{X}$ is *Bochner integrable* with respect to $(\Omega, \mathscr{A}, P)$ if (a) X is Borel ($\equiv$ weakly) measurable and (b) $\| X \|$ is integrable (see [9, Chapter 3]). Under these conditions and the separability of $\mathscr{X}$, the Bochner Integral EX may simply be defined to be the Pettis integral, which here exists and is uniquely determined (see [9, p. 79–80]).

An arbitrary subset W of the product space $\mathscr{Y} \times \mathscr{Y}$ determines a binary relation $\sim^W$ defined as follows: $y \sim^W z \Leftrightarrow (y, z) \in W$. A function $f: C \to \mathscr{Y}$ is *W-convex* on C if

$$f\left(\sum_{k=1}^{n} \alpha_k x_k\right) \overset{W}{\sim} \sum_{k=1}^{n} \alpha_k f(x_k)$$

for all integers $n \geqslant 2$ whenever each $x_k \in C$ and $\alpha_k \geqslant 0$ with $\sum \alpha_k = 1$. (Here we cannot deduce this relation for $n > 2$ from the case $n = 2$.) The function f is *W-continuous* at a point $x \in C$ if the following condition is satisfied: if $\{x_n\} \subset C$ and $\{z_n\} \subset \mathscr{Y}$ are sequences such that $x_n \to x$, $z_n \to z$, and $(f(x_n), z_n) \in W$ for each n, then $(f(x), z) \in W$. (For nonmetrizable $\mathscr{Y}$, we would replace sequences by nets.)

Since Bochner integrability implies Pettis integrability, Theorem 3.1 provides conditions which guarantee that $EX \in C$; these will not be repeated here. The main result concerning W-convex Bochner integrable functions is the following:

THEOREM 4.1. *Under assumptions* B.1–B.7, *if $EX \in C$, and if f is W-continuous at EX, then $f(EX) \sim^W Ef(X)$.*

Proof. Let $X_1, X_2, \ldots$ be a sequence of independent $\mathscr{X}$-valued random vectors, each distributed according to the probability law of X. Since f is W-convex,

$$f\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) \overset{W}{\sim} \frac{1}{n}\sum_{k=1}^{n} f(X_k)$$

for each n. By the SLLN for Bochner-integrable random vectors ([2, 8, 13]),

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to EX \qquad \text{and} \qquad \frac{1}{n}\sum_{k=1}^{n} f(X_k) \to Ef(X) \qquad \text{a.s.}$$

Since $EX \in C$ and f is W-continuous at EX, this implies that $f(EX) \sim^W Ef(X)$. ∎

Remark 4.1. If f is continuous at x, and if W is closed (in the product topology), then f is W-continuous at x. Notice that W is closed iff $\sim^W$ is a closed binary relation in the sense of condition (ii) of the definition of a closed cone ordering (Section 2).

Remark 4.2. If $\sim^W$ is actually a component-wise ordering $\leqslant_T$, and if $t(f(x))$ is lower semicontinuous at x for each $t \in T$, then f is W-continuous at x. Therefore the continuity conditions assumed in Theorems 3.2–3.10 are special cases of W-continuity.

Remark 4.3. The Bochner integral can be defined for mappings assuming values in a separable Frechet space (complete metrizable locally convex LTS) and the SLLN continues to hold (Ahmad [1, Proposition 2, p. 114]). Thus B.1 can be weakened to the assumption that $\mathscr{X}$ and $\mathscr{Y}$ are separable Frechet spaces. Furthermore, the assumption of separability can be replaced by the assumption that the mappings $X: \Omega \to \mathscr{X}$ and $f(X): \Omega \to \mathscr{Y}$ are almost separably valued (see [8] and [9], p. 72).

Acknowledgment

Thanks are due to Professors George Duncan and Roger Wets and the referee for helpful suggestions.

References

[1] Ahmad, S. (1965). Éléments aléatoires dans les espaces vectoriels topologiques. *Ann. Inst. H. Poincaré Sect. B* **2** 95–135.

[2] Beck, A. (1963). On the strong law of large numbers. In *Ergodic Theory* (F. B. Wright, Ed.), pp. 21–53. Academic Press, New York.

[3] Bourbaki, N. (1965). *Éléments de Mathématique, Livre VI*, Intégration (2nd ed. rev.), Chaps. 1–4. Hermann, Paris.

[4] Bourgin, R. D. (1971). Barycenters of measures on certain noncompact convex sets. *Trans. Amer. Math. Soc.* **154** 323–340.

[5] Choquet, G. (1969). *Lectures on Analysis*, Vols. 1–3. W. A. Benjamin, New York.

[6] Ferguson, T. S. (1967). *Mathematical Statistics*. Academic Press, New York.

[7] Groves, T. and Rothenberg, T. (1969). A note on the expected value of an inverse matrix. *Biometrika* **56** 690–691.

[8] Hanš, O. (1957). Generalized random variables. *Trans. First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pp. 61–104. Prague.

[9] Hille, E. and Phillips, R. S. (1957). *Functional Analysis and Semi-Groups* (rev. ed.). Amer. Math. Soc. Colloquium Publication Vol. 31. Providence, Rhode Island.

[10] Khurana, S. S. (1969). Measures and barycenters of measures on convex sets in locally convex spaces. I. *J. Math. Anal. Appl.* **27** 103–115.

[11] Loeve, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton, NJ.

[12] Meyer, P. A. (1966). *Probability and Potentials*. Blaisdell, Waltham, MA.

[13] Mourier, E. (1953). Éléments aléatoires dans un espace de Banach. *Ann. Inst. H. Poincaré* **13** 161–224.

[14] Pettis, B. (1938). On integration in vector spaces. *Trans. Amer. Math. Soc.* **44** 277–304.

[15] Rudin, W. (1973). *Functional Analysis*. McGraw–Hill, New York.

[16] Srivastava, V. K. (1970). On the expectation of the inverse of a matrix. *Sankhyā Ser. A* **32** 236.